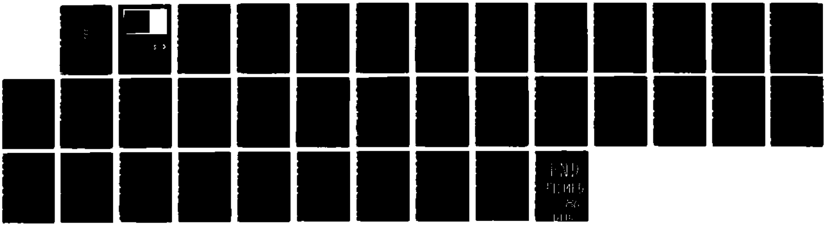
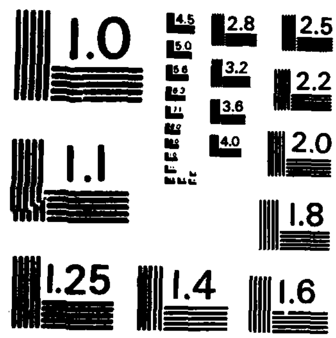
AD-A163 599 THE ORDINARY LEAST SQUARES ESTIMATION FOR THE GENERAL-LINK LINEAR MODELS W..(U) WISCOOSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER N DUAN ET AL. 1/1

UNCLASSIFIED OCT 85 MRC-TSR-2880 DAAG29-80-C-0041 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

MRC Technical Summary Report #2880

THE ORDINARY LEAST SQUARES ESTIMATION FOR THE GENERAL-LINK LINEAR MODELS, WITH APPLICATIONS

Naihua Duan and Ker-Chau Li

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

October 1985

(Received June 17, 1985)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

THE ORDINARY LEAST SQUARES ESTIMATION FOR THE GENERAL-LINK LINEAR MODELS, WITH APPLICATIONS

Naihua Duan* and Ker-Chau Li**

Technical Summary Report #2880
October 1985

ABSTRACT

For a general link linear model (GLLM), we show that the OLS estimate of the slope vector is strongly consistent up to a multiplicative scale, even though the model might actually be nonlinear. Furthermore, the estimated slope vector is strongly consistent for the average slope vector, the average of the pointwise slope vectors on the response surface. For a GLLM with a completely specified link function, we can solve for the multiplicative scalar and estimate the true slope vector, and estimate the intercept and Cox and Snell's generalized residuals. We then estimate the response surface and the pointwise slopes using a generalization of the smearing estimate in Duan (1983). The results can be applied to a number of important subclasses of GLLM, including general transformation models, general scaled transformation models, generalized linear models, dichotomous regression, and Tobit regression.

AMS (MOS) Subject Classifications: 62-XX, 62G05

Key Words: general link function, spherical symmetry, polynomial normality, nonparametric regression, smearing estimate, linearized response surface, average slope vector, dichotomous regression, Tobit regression

Work Unit Number 4 - Statistics and Probability

---

* Statistician at the Rand Corporation, 1700 Main Street, Santa Monica, CA, U.S.A. Part of this research was done while the author was visiting the Mathematics Research Center, University of Wisconsin-Madison. Part of the research was done at the Rand Corporation, supported in parted by a cooperative agreement with SIMS and U.S. EPA, and in part by Rand Corporate research fund.

** Associate Professor at UCLA, Los Angeles, CA, U.S.A. Part of this research was done while the author was visiting the Mathematics Research Center, University of Wisconsin-Madison.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. Part of the research was supported by the National Science Foundation under Grant No. MCS 83-01587.

-i

SIGNIFICANCE AND EXPLANATION

We justify the OLS estimation for regression models on a fundamental level: the method gives meaningful results even when the model is grossly misspecified. The OLS estimates are easy to implement, and can serve as initial estimates for iterative algorithm to derive efficient estimates.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

THE ORDINARY LEAST SQUARES ESTIMATION
FOR THE GENERAL-LINK LINEAR MODELS, WITH APPLICATIONS

Naihua Duan* and Ker-Chau Li**

1. INTRODUCTION

1.1 General-Link Linear Model (GLLM)

We consider a general class of regression models which relate the dependent variable y to the regression (independent) variables x, consider as a row vector:

$$y_i = g(\beta_0 + x_i\beta, \varepsilon_i), \quad i = 1,\ldots,n \quad . \tag{1.1}$$

We will call those models the <u>general-link linear models</u> (GLLM), the function g the <u>general link function</u> (GLF), the parameter vector β the <u>slope vector</u>, and the linear combination $x_i\beta$ the <u>linear component</u>.

Depending on the specific application, the GLF g might be completely unspecified, partially specified, or completely specified. Note that if the GLF is completely unspecified, the intercept β_0 is unidentified, the slope vector β is identified only up to a multiplicative scale:

<u>Observation 1</u>. For any location and scale adjustments on $\beta_0 + x_i\beta$, we can always find a GLF which satisfy the following:

$$g[a + b(\beta_0 + x_i\beta), \varepsilon_i] = g^*(\beta_0 + x_i\beta, \varepsilon_i) \tag{1.2}$$

for any a and b. ∎

* Statistician at the Rand Corporation, 1700 Main Street, Santa Monica, CA, U.S.A. Part of this research was done while the author was visiting the Mathematics Research Center, University of Wisconsin-Madison. Part of this research was done at the Rand Corporation, supported in part by a cooperative agreement with SIMS and U.S. EPA, and in part by Rand Corporate research fund.

** Associate Professor at UCLA, Los Angeles, CA, U.S.A. Part of this research was done while the author was visiting the Mathematics Research Center, University of Wisconsin-Madison.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. Part of the research was supported by the National Science Foundation under Grant No. MCS 83-01587.

When the GLF is partially specified, observation 1 may or may not hold. We will call a subclass of GLLM's identified (unidentified) in location and scale depending on whether observation 1 is violated or satisfied for the subclass.

The GLLM includes many important classes of regression models as special cases. Some important examples are given below. (For all the examples, we assume that the error terms ε_i are identically and independently distributed according to $F(\varepsilon)$.)

Example 1.1. General transformation models. Assume the GLF has the following form:

$$y_i = g(\beta_0 + x_i\beta + \varepsilon_i) \quad . \tag{1.3}$$

We will call this calss of GLLM's the general transformation models. If the GLF is invertible, we can transform y into a linear model:

$$g^{-1}(y_i) = \beta_0 + x_i\beta + \varepsilon_i \quad ,$$

which is the usual specification for transformation models. However, we do not require the GLF g to be invertible.

Example 1.2. Dichotomous regression models. If the GLF in the general transformation model (1.2) is dichotomous,

$$g(t) = 1 \quad \text{if} \quad t > 0 \quad ,$$

$$g(t) = 0 \quad \text{if} \quad t < 0 \quad ,$$

we have a dichotomous regression model. If we further assume that the error distribution $F(\varepsilon)$ is logistic (normal), we have the logistic (probit) regression model. Note that the dichotomous GLF is not invertible.

Example 1.3. Tobit regression model. Assume in (1.3) that the GLF is the censoring function:

$$g(t) = t \quad \text{if} \quad t > 0 \quad ,$$

$$g(t) = 0 \quad \text{if} \quad t < 0 \quad ,$$

i.e.,

$$y_i = \max(\beta_0 + x_i\beta + \varepsilon_i, 0) \quad , \tag{1.4}$$

we have the Tobit regression model. (See, e.g., Maddala, 1983, p. 151.) Note that the censoring GLF (1.4) is not invertible.

Example 1.4. Additive-error models. Assume that the GLF is additive, we have

$$y_i = g(\beta_0 + x_i\beta) + \varepsilon_i \quad . \tag{1.5}$$

Example 1.5. Generalized linear models. Assume the following form for the GLF:

$$y_i = \nu(\beta_0 + x_i\beta) + \sigma(\beta_0 + x_i\beta)\cdot\varepsilon_i \quad . \tag{1.6}$$

This is analogous to the generalized linear model with link function ν and variance function σ^2. Note that our use of the term link function is different from McCullagh and Nelder, 1983. The link function $\eta = \eta(\mu)$ in McCullagh and Nelder is the inverse of ν.

Example 1.6. General scaled transformation models. Efron (1983) considers a rich class of transformation models:

$$y_i = g(\nu(\eta_i) + \sigma(\eta_i)\cdot\varepsilon_i) \quad .$$

If we assume a linear model for η, we have

$$y_i = g(\nu(\beta_0 + x_i\beta) + \sigma(\beta_0 + x_i\beta)\cdot\varepsilon_i) \quad ; \tag{1.7}$$

we will follow Efron's terminology and refer to this class of models as the general scaled transformation models. This class includes both the general transformation model (1.3) and the generalized linear models (1.6) as special cases.

Example 1.7. A combined error model. Assume that the latent variable η has a linear model on the logarithmic scale with behavioral error ε_1:

$$\eta_i = \exp(\beta_0 + x_i\beta + \varepsilon_{1i}) \quad ,$$

while the observed dependent variable measures the latent variable with measurement error ε_2:

$$y_i = \exp(\beta_0 + x_i\beta + \varepsilon_{1i}) + \varepsilon_{2i} \quad .$$

This model belongs to the GLLM, but does not belong to any of the subclasses discussed above. (We do not require the error term ε to be one-dimensional.)

Example 1.8. One-parameter families. Consider any one-parameter family of probability distributions:

$$y_i | \theta_i \sim F(y_i; \theta_i) \quad .$$

If we assume a linear model $\theta_i = \beta_0 + x_i\beta$, we have the following GLLM:

$$y_i = F^{-1}(\varepsilon_i; \beta_0 + x_i\beta) \quad ,$$

where $\varepsilon_i \sim U(0,1)$, F^{-1} is the inverse of $F(\cdot;\theta)$ for a fixed θ. For discontinuous distributions, we define $F^{-1}(\varepsilon) = \{y' : \varepsilon < F(y')\}$.

1.2 OLS Estimation in GLLM

We have a general result on the strong consistency of the OLS estimate for the slope vector β in GLLM.

Theorem 1 Consider the OLS slope coefficients

$$b = (X'QX)^{-1}X'QY \tag{1.8}$$

for the GLLM (1.1), where $Q = I - 1(1'1)^{-1}1'$. We have

$$b \to \gamma \cdot \beta \text{ (a.s.)} \quad , \tag{1.9}$$

where

$$\gamma = \mathrm{Cov}(x\beta, y)/\mathrm{Var}(x\beta) \quad , \tag{1.10}$$

under the following assumptions.

(1a) The GLF g is measurable.

(1b) The regressor variables x_i and error terms ε_i are identically and independently distributed according to the c.d.f. $M(x)F(\varepsilon)$. (Thus x and ε are stochastically independent.)

(1c) The moments $E(x) = \mu$, $\mathrm{Cov}(x) = \Sigma$, and $\mathrm{Cov}(x,y)$ exist; Σ is nonsingular.

(1d) The distribution $M(x)$ of the regressor variables x is spherically symmetric centered at μ with respect to the inner product $\langle v,w\rangle = v'\Sigma w$: for any matrix A such that $A'\Sigma A = \Sigma$, the rotated regressor variables $\mu + (x-\mu)A$ has the same distribution $M(x)$ as the original regressor variables x.

The spherical symmetry condition (1d) can be replaced by the following polynomial normality conditions.

(1e) The GLF g is a polynomial in $x\beta$ of degree k.

(1f) The regressor variables can be orthonormalized to be stochastically independent; i.e., there exist a square matrix A such that $A\Sigma A' = I$, and the orthonormalized regressor variables $\{(x-\mu)A^j\}$ are mutually independent.

(1g) The first $k + 1$ cumulants of all the othonormalized regressor variables are identical to those of the standard normal variate. ∎

(Proof) By the strong law of large numbers and conditions (1b), (1c), we have

$$b \to \Sigma^{-1}\Sigma_{xy}' \quad (a.s.) \quad ,$$

where $\Sigma_{xy} = Cov(x,y)$. The rest of the proof is given in section two. ∎

__Remark 1.1__. The result in theorem 1 does not apply to the intercept β_0; it also leaves the multiplicative scalar γ to be determined. For completely unspecified GLLM's, this is the best that can be achieved - according to observation 1, the intercept β_0 and the multiplicative scalar γ are not identified and can be absorbed into the GLF g. The same comment applies to partially specified GLLM's which are unidentified in location and scale, i.e., for which observation 1 is valid.

For subclasses of GLLM unidentified in location and scale, we will consider the estimation of the response surface $E(y|x)$ in section 1.4. Note that while the intercept β_0 and the multiplicative scalar γ are not identified, the response surface is invariant to location and scale adjustments in the linear component $x\beta$.

For a subclass of GLLM identified in location and scale, it is usually possible to estimate the intercept β_0 and the multiplicative scalar γ. We give the results in section 3 for GLLM's with completely specified GLF's.

__Remark 1.2__. It is well known that the OLS estimate for linear models is robust to perturbations in the error distribution in the sense that under assumption (1b) and (1c), the OLS estimate is strongly consistent. (Note that for a linear model assumption (1c) implies that $E(\varepsilon)$ exists; since we are not concerned with estimating the intercept in

theorem 1, we do not require $E(\varepsilon) = 0$.) Theorem 1 indicates that the OLS estimate for GLLM is robust both to perturbations in the GLF as well as to perturbations in the error distribution in the sense that asymptotically the OLS gives the correct answer in the sense of (1.9).

Remark 1.3. Beyond strong consistency, we usually like to find efficient or nearly efficient estimates if the model is sufficiently specified. The OLS estimate, despite its strong consistency, is unlikely to be efficient except in some special cases. In a sufficiently specified model, we might be able to find more efficient estimators, e.g., using adaptive methods. In those situations, it is usually desirable to use a consistent estimate as the starting value for numerical algorithms aimed at finding efficient estimates. Therefore theorem 1 indicates that the OLS estimates (up to a multiplicative scalar) can be used as the starting value. In a completely specified model, it is usually posssible to determine or estimate the multiplicative scalar. (See section 3.)

Researchers at the Rand Corporation have been using the OLS estimates to derive starting values for logistic regression models since 1974 and for probit regression models since 1980. Generally speaking the results are very satisfactory. For both models, a stringent convergence criterion that the conditional log-likelihood move by no more than 0.01 is satified within two or three iterations.

Remark 1.4. If the multiplicative scalar γ is zero, there is no information in y for β on the linear scale; we might still be able to estimate β by transforming y to an alternative scale on which γ is nonzero. Throughout this paper will assume γ is nonzero. See remark 1.8 for a testing procedure for this hypothesis for the general transformation model.

Remark 1.5. As a special case of the result based on the polynomial normality conditions, consider a GLF which is quadratic in $x_i\beta$. The strong consistency result (1.8) holds if all orthonormalized regressor variables are symmetric. This will be true, e.g., in a

factorial experiment.

Remark 1.6. Theorem 1 requires fairly strong assumptions on the regressor variables, namely, they have to be spherically symmetric (1d) or polynomially normal (1e - 1g). This should be considered as a desirable property for experimental designs, namely, designs which satisfy either condition (1d) or (1e - 1g) are robust in the sense of remark 1.2. For example, spherically symmetric designs enjoy the robustness property under theorem 1. As another example, we might known from prior experience or theory that the linear model is applicable on the cube root scale but then we have to retransform back to the original scale. A design in which all factors are symmetric and also has zero kurtosis is the robust in the sense of remark 1.2.

Remark 1.7. Conditions (1d) or (1e - 1g) rule out the possibility of having interaction terms in the linear components $x\beta$. In other words we need to assume that under the appropriate link function, the effects of the regressor variables are additive. For example, for the general transformation model, this amounts to the assumption that after an appropriate transformation we have *an additive model*. See Scheffé, 1959, pp. 95-98, for a characterization of response surfaces which can be linearized into an additive model.

1.3 Response Surface and Linearized Response Surface.

In many situations our ultimate goal is to estimate the response surface

$$\nu(x) = \tilde{\nu}(x\beta) = E(y|x) = \int g(\beta_0 + x\beta,\varepsilon)dF(\varepsilon) \quad . \tag{1.11}$$

Note that the response surface depends on the regressor variables x only through $x\beta$. If we know the true parameters β, we can estimate the response surface using a nonparametric regression of y_i on $x_i\beta$. For the completely unspecified GLLM or a subclas which is unidentified in location and scale, the indeterminancy of the intercept β_0 and the multiplicative scalar γ is irrelevant. For any two GLF's g and g^* which satisfy (1.2), the response surfaces are the same.

The deviations from the response surface, $y_i - \nu(x_i)$, is not usually homoscedastic. If we know the form of the variance function $\sigma^2(x\beta) = \mathrm{Var}(y|x)$, e.g., if we have an additive-error model or a generalized linear model, we could use weighted methods in the nonparametric regression. The variance function is also a function of the linear component $x_i\beta$. When the form of the variance function is known, we can use iterated weighted nonparametric regression, estimating the variance function from the deviations from the estimated response surface.

In reality we don't know the true parameters β, and have to use an estimate b for it, such as the OLS estimate (1.8). The estimate for the response surface has to be based on the nonparametric regression of y on xb. Therefore it is necessary to consider the error in the regressor variable xb. We don't know of any work in the literature on the consideration of nonparametric regression in the presence of error-in-variable.

For a general transformation model with a completely specified GLF, if the GLF g is invertible, and the OLS estimation is applied on the linear model scale $g^{-1}(y)$, Duan (1983) proposed the smearing estimate as an estimate for the response surface:

$$s(x) = n^{-1} \Sigma_i \, g(\hat{\beta}_0 + x\hat{\beta} + \hat{\varepsilon}_i) \quad ,$$

where $\hat{\beta}_0$, $\hat{\beta}$, and $\hat{\varepsilon}_i$ are OLS estimates based on the regression of $g^{-1}(y)$ on x. Duan (1983) showed that the smearing estimate is weakly consistent for the response surface $\nu(x)$ under some regularity conditions.

For a GLLM with a completely specified GLF, we can use the following generalization of the smearing estimate to estimate the response surface:

$$s(x) = n^{-1} \Sigma_i \, g(b_0 + c^{-1} \cdot xb, e_i) \quad , \tag{1.12}$$

where b is the OLS estimate (1.8); b_0, c, and e_i are estimates for β_0, γ, and ε_i. In section 4 we give a consistency result for the smearing estimate (1.12).

The construction of the estimated response surface, either using nonparametric regression or the smearing estimate, can be computationally expensive. As an alternative, we may consider the linearized response surface

$$\rho(x) = \rho_0 + (x-\mu)\theta \quad , \tag{1.13}$$

where $\rho_0 = E(y)$, $\theta = \Sigma^{-1} \Sigma_{xy}'$, $\Sigma_{xy} = \mathrm{Cov}(x,y)$. The linearized response surface minimizes

the mean squared prediction error $E(y - \alpha_0 - x\alpha)^2$ over all linear surfaces $\alpha_0 + x\alpha$; it also minimizes the mean squared approximation error $E(\nu(x) - \alpha_0 - x\alpha)^2$. The proof in section 1.2 for theorem 1 shows that the OLS estimate b in (1.8) is strongly consistent for $\theta = \Sigma^{-1}\Sigma_{xy}'$ under assumptions (1b) and (1c). Furthermore, the OLS prediction

$$r(x) = \bar{y} + (x - \bar{x})b$$

is strongly consistent for the linearized response surface. What remains to be shown for theorem 1 is that the slope θ in the linearized response surface (1.13) is related to the parameter β in GLLM as follows:

$$\theta = \Sigma^{-1}\Sigma_{xy}' = \gamma\cdot\beta \quad , \tag{1.14}$$

where γ is given in (1.10). We will prove this result in section 2.

<u>Corollary 1</u>. Under the assumptions in theorem 1, the linearized response surface has the following expression:

$$\rho(x) = \rho_0 + \gamma\cdot(x-\mu)\beta \quad ,$$

where γ is given in (1.9). In particular, the linearized response surface depends on the *regressor variables* x *only through* $x\beta$. ∎

1.4 Pointwise Slopes and Average Slopes

In many situations we are not interested in the response surface $\nu(x)$ per se. Instead we are interested in the pointwise slopes, assuming that they exist:

$$\nabla_x\nu(x) = \beta\cdot\tilde{\nu}'(x\beta) \quad . \tag{1.15}$$

For a fixed design point x, the pointwise slope $\nabla_x\nu(x)$ is the change in the mean response when the levels of the regressor variables change. Note that in GLLM the pointwise slopes are always proportional to β, therefore the OLS b in (1.10) is strongly consistent for all pointwise slopes up to the multiplicative scale $\gamma/\tilde{\nu}'(x\beta)$.

The multiplicative scalar $\tilde{\nu}'(x\beta)$ in the pointwise slopes (1.15) can be estimated by differentiating the estimated response surface, either the nonparametric regression estimate for ν, or the smearing estimate s in (1.12). We will discuss the latter

method in Section 4.

The multiplicative scalar $\tilde{\nu}'(x\beta)$ is given by

$$\tilde{\nu}'(x\beta) = \frac{d}{d(x\beta)} \int g(\beta_0 + x\beta, \varepsilon)dF(\varepsilon) \quad . \tag{1.16}$$

Assuming that g is differentiable, and that the differentiation and integration can be interchanged, we have

$$\tilde{\nu}'(x\beta) = \int g_1(\beta_0 + x\beta, \varepsilon)dF(\varepsilon) = E[g_1|x] \quad , \tag{1.17}$$

where g_1 is the first partial derivative of g.

The same comment in section 1.3 regarding the cost of estimating the response surface $\nu(x)$ can also be applied to the estimation of the pointwise slopes $\beta\cdot\tilde{\nu}'(x\beta)$. As an alternative, we can instead estimate the average slopes

$$E_x\nabla_x\nu(x) = \beta\cdot E_x\tilde{\nu}'(x\beta) \quad . \tag{1.18}$$

If the expression (1.17) is valid, the average slopes are given by

$$E_x\nabla_x\nu(x) = \beta\cdot E\, g_1(\beta_0 + x\beta, \varepsilon) \quad . \tag{1.19}$$

Expression (1.19) is usually easier to evaluate than expression (1.18); we don't need to evaluate the conditional expectation $\nu(x) = E(y|x)$. However, there are important models such as logistic regression for which the differentiation and integration in (1.16) cannot be interchanged.

Stein (1981) gave the following lemma which was a crucial tool for evaluating MSE for the estimation of the normal mean.

<u>**Stein's Lemma**</u>. Let u be a $N(\mu,\sigma^2)$ real random variables and let the real-valued function $g(u)$ be the indefinite integral of the Lebesque measurable function $g'(u)$. Assume that $E|g'(u)| < \infty$. Then

$$E[g'(u)] = \mathrm{Cov}(u,g(u))/\mathrm{Var}(u) \quad . \quad \blacksquare$$

We will follow Stein's terminology and refer to g as an <u>almost differentiable</u> function, and refer to g' as the derivative of g. Note that the almost differentiable condition is stronger than being differentiable almost everywhere.

Theorem 2. The multiplicative scalar (1.10) in theorem 1 is identical to the multiplicative scalar in the average slopes (1.18),

$$Cov(x\beta,y)/Var(x\beta) = E_x \tilde{\nu}'(x\beta) \quad , \tag{1.20}$$

under assumptions (1a - 1c) in theorem 1 and the following assumptions.

(2a) The response surface $\tilde{\nu}(x\beta)$ is almost differentiable in $x\beta$.

(2b) $x\beta$ is normally distributed.

The normality assumption (2a) can be replaced by the following polynomial normality assumptions analogous to (1e - 1g).

(2c) The response surface $\tilde{\nu}(x\beta)$ is polynomial in $x\beta$ of degree k.

(2d) The first $k + 1$ cumulants of $x\beta$ are identical to those of the standard normal variate.

The expression (1.19) can be used in (1.20):

$$Cov(x\beta,y)/Var(x\beta) = E\ g_1(\beta_0 + x\beta,\varepsilon) \quad , \tag{1.21}$$

under the assumptions (1a) - (1c) in theorem 1, assumption (2b), and the following additional assumptions.

(2e) For all ε, the GLF $g(\eta,\varepsilon)$ is almost differentiable with respect to η.

(2f) The differentiation and integration can be interchanges in *(1.16)*. *For example,* this is satisfied if $|g_1|$ is dominated uniformly by an integrable function Q:

$$|g_1(\eta,\varepsilon)| \le Q(\varepsilon),\ E\ Q(\varepsilon) < \infty \quad .$$

Assumption (2b) can be replaced by the following polynomial normality conditions.

(2g) The GLF g is a polynomial in $x\beta$ of degree k.

(2h) The first $k + 1$ cumulants of $x\beta$ are identical to those of the standard normal variable. ‖

(Proof) Applying Stein's Lemma under the assumption that $x\beta$ is normally distributed, we have

$$E_x \tilde{\nu}'(x\beta) = Cov(x\beta,\ \tilde{\nu}(x\beta))/Var(x\beta) \quad .$$

Note that

$$\begin{aligned} Cov(x\beta,y) &= E[(x-\mu)\beta\cdot g(\beta_0 + x\beta,\varepsilon)] \\ &= E\ E^x(x-\mu)\beta\cdot g(\beta_0 + x\beta,\varepsilon) \\ &= E(x-\mu)\beta\cdot\tilde{\nu}(x\beta) \\ &= Cov(x\beta,\tilde{\nu}(x\beta)) \quad . \end{aligned}$$

Note that we adopt Stein's notation tht a superscript for the operators E, Var, and Cov indicates applying the operator conditioned on the superscript.

The result under the polynomial normality conditions (2c) and (2d) are proved in the same way as the proof for theorem 1 under the polynomial normality assumptions (1e - 1g), to be given in section 2. ∎

<u>Remark 1.8</u>. For a general transformation model with a completely specified invertible GLF g, if expression (1.21) is valid, we can estimate the multiplicative scalar $\gamma = Eg'(g^{-1}(y))$ by the sample average $A_n = n^{-1}\Sigma_i g'(g^{-1}(y_i))$, which converges almost surely to γ by the strong law of large numbers. If the second moment for g' exists, $Z_n = n^{1/2}(A_n-\gamma)/s_n$ converges in law to the standard normal distribution, where s_n is the sample standard deviation. We can therefore construct confidence intervals for γ; in particular, we can test the null hypothesis $\gamma = 0$.

<u>Corollary 2</u>. Under the assumptions in theorems 1 and 2, the OLS estimate b in (1.8) has the following almost sure limit:

$$b \to \gamma\cdot\beta \quad (a.s.) \quad ,$$

where
$$\gamma = E_x\tilde{\nu}'(x\beta)$$

or
$$\gamma = E\ g_1(x\beta,\varepsilon) \quad . \quad ∎$$

<u>Corollary 3</u>. Under the assumptions in theorems 1 and 2, the linearized response surface has the alternative expression

$$\rho(x) = \rho_0 + \gamma\cdot(x - \mu)\beta \quad ,$$

where γ is given as in corollary 2. ∎

2. PROOF AND DISCUSSION OF THEOREM 1

In this section we given the proof for theorem 1, and give several useful generalizations.

Proof of theorem 1

We need to prove that

$$v = E(x-\mu)y = \gamma \cdot \beta'\Sigma \quad .$$

Let θ be an p dimensional column vector such that $\beta'\Sigma\theta = 0$, $\theta'\Sigma\theta = 1$. We will show that $v\theta = 0$. Note that

$$\begin{aligned} v\theta &= E(x-\mu)\theta \cdot g(\beta_0 + x\beta, \varepsilon) \\ &= E\, E^{\varepsilon, x\beta}(x-\mu)\theta \cdot g(\beta_0 + x\beta, \varepsilon) \\ &= E[g(\beta_0 + x\beta, \varepsilon) \cdot E^{\varepsilon, x\beta}(x-\mu)\theta] \\ &= E[g(\beta_0 + x\beta, \varepsilon) \cdot E^{x\beta}(x-\mu)\theta] \quad . \end{aligned}$$

Since $\mathrm{Cov}(x\beta, x\theta) = \beta'\Sigma\theta = 0$, $x\beta$ and $x\theta$ are uncorrelated. If x has a multinormal distribution, it follows that $x\beta$ and $x\theta$ are stochastically independent, therefore

$$E^{x\beta}(x-\mu)\theta = 0 \quad , \tag{2.1}$$

thus $v\theta = 0$. We prove in Appendix A that the same condition holds if x is spherically symmetric.

Having proved that $v\theta = 0$ for all θ such that $\beta'\Sigma\theta = 0$, it follows that v must fall along the direction $\beta'\Sigma$, i.e., $v = \alpha \cdot \beta'\Sigma$ for some scalar α. We now determine the scalar α. On the one hand,

$$v\beta = \alpha \cdot \beta'\Sigma\beta = \alpha \cdot \mathrm{Var}(x\beta) \quad .$$

On the other hand,

$$v\beta = E(x-\mu)\beta \cdot y = \mathrm{Cov}(x\beta, y) \quad .$$

Therefore $\alpha = \mathrm{Cov}(x\beta, y)/\mathrm{Var}(x\beta)$ as given in (1.9) in theorem 1.

We now prove the theorem under the polynomial normality conditions (1e - 1g). Let $\xi = (x-\mu)A$ be the orthonormalized regressor variables given in (1f). Note that

$$\beta_0 + x\beta = \beta_0^* + \xi\varphi \quad ,$$

$$v\theta = E(x-\mu)\theta \cdot y \tag{2.2}$$

$$= E\ \xi\tau \cdot g(\beta_0^* + \xi\varphi,\ \epsilon) \quad ,$$

where $\beta_0^* = \beta_0 + \mu\beta$, $\varphi = \Sigma A'\beta$, $\tau = \Sigma A'\theta$.

By assumption (1e), g is a polynomial in $\xi\varphi$ of degree k; it follows that the integrand in (2.2) is a multivariate polynomial in $(\xi_1,\ldots,\xi_p)$ of degree k:

$$v\theta = E\{\Sigma_J[a_J\ \Pi_{\rho=1}^{p}(\xi_\rho)^{j(\rho)}]\} \quad , \tag{2.3}$$

where the summation is taken over all index sets $J = \langle j(1),\ldots,j(p)\rangle$ such that $\Sigma_\rho j(\rho) \le k$. The coefficients a_J depend on Σ, β_0^*, φ, and τ.

Since the orthonormalized regressor variables ξ are assumed to be mutually independent, the expectation in (2.3) can be taken term by term:

$$v\theta = \Sigma_J[a_J\ \Pi_{\rho=1}^{p}\ m_{\rho,j(\rho)}] \quad , \tag{2.4}$$

where $m_{\rho,j(\rho)}$ is the $j(\rho)$-th moment for ξ_ρ.

By assumption (1g), the moments m are identical to those of a corresponding standard normal variate. Therefore $v\theta$ has the same value if we replace ξ by the corresponding multinormal random vector $\xi^* \sim N(0,I)$. Since the multinormal random vector is spherically symmetric, by the first part of theorem 1, we have $v\theta = 0$. ∎

Remark 2.1. For logistic regression, a special case of example 1.2, the result in theorem 1 for normally distributed regressor variables has been known since Fisher (1936). See Haggstrom (1983) for a comprehensive discussion of the OLS estimation for the logistic regression model. As was noted in section 1.2, Rand researchers have been using the OLS estimates to derive starting values for both the logistic and probit regressions, with generally satisfactory results. For probit regression, this is motivated by the fact that the normal c.d.f. is very well approximated by the logistic function after the appropriate scale adjustment. See Haggstrom (1983, p. 236) for a brief discussion. Theorem 1 gives further justification for the use of OLS estimates to devise starting values probit regression.

Remark 2.2. Goldberger (1981) showed a result similr to theorem 1 for linear models after selection under the assumption that the regressor variables x and the error terms ε are jointly normal. (See also Maddala 1983, pp. 168-170.) Chung and Goldberger (1984) showed a similar result under the assumption that $E(x|y)$ is linear.

Remark 2.3. Brillinger (1982) showed the result in theorem 1 for the additive-error model in example 1.4, under the assumption that the regressor variables x are normally distributed. Brillinger also gave expression (1.21) in theorem 2 for the normal case. In an example, Brillinger also commented on the generalization of the theorem to deal with the Tobit regression model in example 1.3. For consideration of experimental designs, the spherical symmetry condition in theorem 1 is an important improvement over the restriction to multinormal distributions.

Remark 2.4. It follows from Cacoullous' (1967) theorem 1 that the only spherically symmetric distribution whose components are mutually independent is the multinormal distribution. Otherwise the components are stochastically dependent, even though they are uncorrelated. See, also, Kagan et al (1973, chapter 5). Therefore the only intersection between the spherical symmetry condition (1d) and the polynomial symmetry condition (1e - 1g) is the normal case.

Remark 2.5. In the proof of theorem 1, we do not need to consider the entire slope vector β simultaneously.

Corollary 4. Let the regressor variables x be partitioned into two parts, x_1 and x_2; partition the slope vector β and the OLS estimate b correspondingly into β_1, β_2 and b_1, b_2. We have

$$b_1 \rightarrow \gamma_1 \cdot \beta_1 \quad (a.s) \quad , \tag{2.5}$$

where

$$\gamma_1 = \mathrm{Cov}(x_1\beta_1,\ y)/\mathrm{Var}(x_1\beta_1) \quad , \tag{2.6}$$

under assumptions (1a - 1c) in theorem 1, and the following additional assumptions.

(1h) The two subsets of regressor variables x_1 and x_2 are stochastically

independent.

(1i) The subset of regressor variables x_1 are either spherically symmetric (1d) or polynomially normal (1e - 1g). ∎

The multiplicative scalar γ_1 can be expressed alternatively as

$$\gamma_1 = E\tilde{v}'(\beta_0 + x\beta,\ \varepsilon) \quad , \tag{2.7}$$

or

$$\gamma_1 = E\ g_1(\beta_0 + x\beta,\ \varepsilon) \quad , \tag{2.8}$$

under the assumption in theorem 2, with $x_1\beta_1$ replacing $x\beta$ in (2b), (2d), and (2h). ∎

(Proof). In the proof for either theorem, replace E^{ε} by E^{ε,x_2}. ∎

Remark 2.6. Corollary 4 allows us to consider the slope parameters in subsets. In the minimal, if we have at least one symmetrically distributed regressor variable, corollary 4 can be applied. (A symmetric variable is spherically symmetric!)

Remark 2.7. When assumption (1h) is not satisfied, we might still be able to orthogonalize the two subsets of regressor variables and apply Corollary 3. Consider, for example, an observational study in which we draw random samples from two different populations; let x_1 be the label for the population, x_2 be the regressor variables. Assume that the distribution of x_2 in the two populations are different by a location shift:

$$x_2 = x_1\mu + z_2 \quad ,$$

where z_2 is independent of x_1; assume that z_2 satisfies the conditions in corollary 4. If we know the shift μ, we can reparametrize the model as follows:

$$y = g(\beta_0 + x_1(\beta_1 + \mu\beta_2) + (x_2 - x_1\mu)\beta_2,\ \varepsilon) \quad ,$$

and apply corollary 2 to β_2. In reality, we need to estimate μ from the sample.

Remark 2.8. For the additive-error model in example 1.4, we don't need the error terms ε_i to be identically distributed; they don't even need to be independent of the regressor variables x_i. All that is required is that when we apply OLS, the term corresponding to

ε converge to zero:

$$n^{-1}\Sigma_i(x_i - \bar{x})'\varepsilon_i \to 0 \quad \text{(a.s.)} \quad . \tag{2.9}$$

In a sense, all GLLM's can be considered as additive error models:

$$y = g(\beta_0 + x\beta, \varepsilon) = \nu(\beta_0 + x\beta) + \varepsilon^* \quad ,$$

where
$$\varepsilon^* = y - \nu(\beta_0 + x\beta) \quad .$$

The distribution of ε^* will depend on x in general. Since conditions that guarantee the convergence in (2.9) for ε^* are not easy to formulate, we will continue to use the GLLM specification.

3. PARAMETER ESTIMATION FOR A COMPLETELY SPECIFIED GLF

When the GLF in a GLLM is completely specified, we can usually estimate the parameters β_0 and γ left undetermined in theorem 1. We can also estimate the residuals ε_i. Throughout this section, we will assume the conditions in theorem 2 for expression (1.21) for the multiplicative scalar γ, i.e., we assume conditions (1a - 1c), (2b), and (2e - 2f).

3.1 GLLM

First we consider the estimation for β_0 and γ in a GLLM with a completely specified GLF. Assume that expression (1.21) in theorem 2 is valid,

$$\gamma = E\, g_1(\beta_0 + x\beta, \varepsilon) \quad . \tag{3.1}$$

Furthermore, we assume

$$E\, \varepsilon = 0 \quad . \tag{3.2}$$

We have

$$b \to \gamma \cdot \beta \quad , \tag{3.3}$$

and

$$y_i = g(\beta_0 + x_i\beta, \varepsilon_i) \quad . \tag{3.4}$$

Replacing the unknown parameters by the corresponding estimates in (3.1), (3.2), and (3.4), we have the following system of simultaneous equations:

$$c = n^{-1}\Sigma_i\, g_1(b_0 + c^{-1} \cdot x_i b, e_i) \quad , \tag{3.5}$$

$$0 = n^{-1}\Sigma_i e_i \quad , \tag{3.6}$$

$$y_i = g(b_0 + c^{-1} \cdot x_i b, e_i) \quad , \tag{3.7}$$

where b is the OLS estimate (1.8), c is our estimate for γ, b_0 estimates β_0, and e_i estimates ε_i. We have $n + 2$ equations with $n + 2$ unknowns, therefore the system (3.5) - (3.7) is solvable, although the solution set might be nonexistent or nonunique. Given that g is a nonlinear function, the equations have to be solved by iteration; given the dimension of the system, they are probably expensive to solve.

The system of equations (3.5) - (3.7) is much easier to solve if the given GLF is invertible in its second variable. In this case we can solve for e_i in (3.7) by

$$e_i = g^{-1}(b_0 + c^{-1} \cdot x_i b, y_i) \quad , \tag{3.8}$$

where we use g^{-1} to denote the inverse of the function $g(\eta, e)$ for a fixed η. Substituting (3.8) into (3.5) and (3.6), we have

$$c = n^{-1}\Sigma_i \; g_1(b_0 + c^{-1}\cdot x_i b, \; g^{-1}(b_0 + c^{-1}\cdot x_i b, \; y_i)) \quad , \tag{3.9}$$

and
$$0 = \Sigma_i \; g^{-1}(b_0 + c^{-1}\cdot x_i b, \; y_i) \quad . \tag{3.10}$$

We need to solve for b_0 and c in the two-equation system (3.9) and (3.10). If both g_1 and g^{-1} are easy to evaluate, it is not difficult to solve this system. Whether the solutions, if any, are consistent needs to be determined for specific GLLM's.

Remark 3.1. The estimated terms e_i in (3.8) are the generalized residuals considered in Cox and Snell (1968).

3.2 General Transformation Models

For a general transformation model (example 1.1) with a completely specified GLF, we can estimate the multiplicative scalar by the sample mean

$$c = n^{-1}\Sigma_i \; g'(g^{-1}(y_i)) + \gamma = Eg'(y) \quad , \tag{3.11}$$

assuming that expression (1.23) in theorem 2 is valid, and that the given link function g is invertible. We can solve for e_i in (4.7) by

$$e_i = g^{-1}(y_i) - b_0 - c^{-1}\cdot x_i b \quad . \tag{3.12}$$

It follows then from (4.6) that

$$b_0 = n^{-1}\Sigma_i \; g^{-1}(y_i) - c^{-1}\cdot \bar{x}b \quad . \tag{3.13}$$

Note that

$$b_0 + E\, g^{-1}(\eta) - \mu\beta = \beta_0 \quad (\text{a.s.}) \quad . \tag{3.14}$$

Example 3.1. For the Tobit regression model in example 1.3, we have

$$\gamma = E\, g'(y) = P(y > 0) \quad .$$

Note that for $y > 0$, $g^{-1}(y) = y$; for $y = 0$, i.e., $\beta_0 + x\beta + \varepsilon < 0$, $g^{-1}(y)$ is undetermined but $g'(\beta_0 + x\beta + \varepsilon) \equiv 0$.

We can therefore estimate γ by

$$c = n^{-1} \cdot \#(y_i > 0) \quad .$$

Note that the dominated convergence condition (2f) in theorem 2 is satisfied for the Tobit

model.

3.3. Generalized Linear Models

For th generalized linear model in example 1.5, expression (1.20) and (1.21) for the multiplicative scalar γ are the same, i.e., we need only consider $\tilde{\nu}'$ in the partial derivative g_1:

$$c = n^{-1}\Sigma_i \tilde{\nu}'(b_0 + c^{-1}x_i b) \quad . \tag{3.15}$$

Furthermore, the error terms can be estimated by

$$e_i = [y_i - \tilde{\nu}(b_0 + c^{-1}\cdot x_i b)]/\sigma(b_0 + c^{-1}\cdot x_i b) \quad , \tag{3.16}$$

and equation (3.6) reduces to

$$\bar{y} = n^{-1}\Sigma_i \tilde{\nu}(b_0 + c^{-1}\cdot x_i b) \quad . \tag{3.17}$$

The estimates b_0 and c can be solved from (3.15) and (3.17).

Example 3.2. Consider a generalized linear model with an exponential link function $\tilde{\nu}(\eta) = \exp(\eta)$. It follows from (3.15) and (3.17) that

$$c = \bar{y} \quad . \tag{3.18}$$

Note that c converges almost surely to $E(y) = E\ \tilde{\nu}(\beta_0 + x\beta) = E\ \tilde{\nu}'(\beta_0 + x\beta) = \gamma$.

Substituting (3.18) in (3.15), we have

$$\exp(b_0) = \bar{y}/[n^{-1}\Sigma_i \exp(x_i b/\bar{y})] \quad . \tag{3.19}$$

The denominator in (3.19) converges almost surely to $E \exp(x\beta)$, therefore $\exp(b_0)$ converges almost surely to $\exp(\beta_0)$.

3.4. General Scaled Transformation Models

For a general scaled transformation model (example 1.6) with a completely specified GLF, we can estimate γ and β_0 by solving the following two equation system:

$$c = n^{-1}\Sigma_i g'(g^{-1}(y_i))\cdot\tilde{\nu}'(b_0 + c^{-1}\cdot x_i b) \quad , \tag{3.20}$$

and

$$n^{-1}\Sigma_i g^{-1}(y_i) = n^{-1}\Sigma_i \tilde{\nu}(b_0 + c^{-1}x_i b) \quad , \tag{3.21}$$

assuming that the transformation function is invertible. The error terms can be estimated

by

$$e_i = [g^{-1}(y_i) - \nu_i]/\sigma_i \quad , \tag{3.22}$$

where $\nu_i = \tilde{\nu}(b_0 + c^{-1}x_i b)$, $\sigma_i = \sigma(b_0 + c^{-1}x_i b)$.

4. SMEARING ESTIMATION FOR THE RESPONSE SURFACE

In section 1.3 we discussed the estimation of the response surface $\nu(x) = E(y|x)$ for a completely unspecified GLLM and for partially specified GLLM's, using nonparametric regression of y_i on $x_i b$. For a GLLM with completely specified GLF, it is still possible to use nonparametric regression to estimate the response surface; however, we should be able to do better by making use of the information available on the GLF. In this section we consider the smearing estimate (1.12) in section 1.3 and derive some of it properties.

4.1. Smearing Estimate for GLLM

For a given GLLM of a general form, the smearing estimate is given by (1.12), with estimates b_0, c, and e_i given in Section 3.1. In section 4.2 we give a consistency result for this estimate.

For a given general transformation model, the smearing estimate is given by

$$s(x) = n^{-1}\Sigma_i\ g(c^{-1}\cdot xb + g^{-1}(y_i) - c^{-1}\cdot x_i b) \quad , \tag{4.1}$$

with c given by (3.11). We give a consistency result for this estimate in section 4.3.

For a given generalized linear model, there is no need to "smearing" over the residuals to estimate the response surface; we can simply use

$$s(x) = \nu(b_0 + c^{-1}\cdot xb) \quad , \tag{4.2}$$

with the estimates b_0 and c given in section 3.3. If the estimates b_0 and c are consistent, (e.g., if the link function $\tilde{\nu}$ is exponential as in example 3.2), and $\tilde{\nu}$ is continuous at $\beta_0 + x\beta$, the estimated response surface is consistent.

The smearing estimate for the general scaled transformation model is

$$s(x) = n^{-1}\Sigma_i\ g(\nu + \sigma\cdot e_i) \quad , \tag{4.3}$$

with $\nu = \nu(b_0 + c^{-1}xb)$, $\sigma = \sigma(b_0 + c^{-1}xb)$; b_0, c, and e_i are given in section 3.4.

4.2. Smearing Estimate for GLLM

If b_0, b, c, and e_i are good estimates for the corresponding unknown quantities, we have

$$s(x) = n^{-1}\Sigma_i\ g(b_0 + c^{-1}xb,\ e_i)$$

$$\approx n^{-1}\Sigma_i\ g(\beta_0 + x\beta,\ \varepsilon_i)$$

$$\to \tilde{\nu}(x\beta) \quad \text{(a.s.)} \quad .$$

Theorem 3. The smearing estimate (1.12) is weakly consistent,

$$s(x) \to \tilde{\nu}(x\beta) \qquad (p) \quad , \tag{4.4}$$

under the assumptions in theorem 1 and the following assumptions.

(3a) The estimates b_0 and c given in section 3.1 exist and are weakly consistent for β_0 and γ.

(3b) The estimates e_i given in section 3.1 exist and satisfy the following condition:

$$\Sigma_i(e_i - \varepsilon_i)^2 = O_p(1) \quad . \tag{4.5}$$

(3c) The GLF $g(\eta,\varepsilon)$ is continuously differentiable jointly in η and ε.

(3d) The following expectations exist for all $M > 0$ and some $a > 0$:

$$E\{\sup[|g_1(\beta_0+x_0\beta+t,\varepsilon+s)|;\ |s| < M,\ |t| < a\} \quad , \tag{4.6}$$

and

$$E\{\sup[|g_2(\beta_0+x_0\beta+t,\varepsilon+s)|^2;\ |s| < M,\ |t| < a\} \quad , \tag{4.7}$$

where for $j = 1,2$, g_j = the j-th partial derivative of g. ∎

(The proof is essentially the same as in Duan (1983) and is given in Appendix B.)

4.3. Smearing Estimate for GTM

We will verify the assumptions in theorem 3 to prove the consistency of the smearing estimate for the general transformation model. First we will prove a result which is required to prove assumption (3b) in theorem 3.

Theorem 4. The OLS estimate b in (1.7) is consistent for $\gamma \cdot \beta$ of order $n^{1/2}$ for the general transformation model:

$$n^{1/2}(b - \gamma \cdot \beta) = O_p(1) \quad , \tag{4.8}$$

under assumptions (1a - c) in theorem 1 and the following assumption:

(4a) The expectation $E\, y^2 \cdot x'x$ exists. ∎

(The proof is given in Appendix C.)

Corollary 5. The smearing estimate (5.1) for the general transformation family is weakly consistent under the assumptions in theorem 1, assumption (4a) in theorem 4, and the following assumptions:

(4e) The GLF g is continuously differentiable.

(4f) The following expectation exist for all $M > 0$:

$$E\{\sup[|g'(\beta_0 + x_0\beta + \varepsilon + t)|^2 ; \ |t| < M]\} \quad . \quad ∎ \tag{4.9}$$

(Proof) It was noted in section 3.2 that the estimates b_0 and c exist and are strongly consistent. The estimates e_i exist and the squared sum in (4.5) can be expressed as follows:

$$\Sigma_i (e_i - \varepsilon_i)^2 = n\bar{\varepsilon}^{\,2} + \Sigma_i [(x_i - \bar{x})(b - c\beta)]^2 / c^2 \quad . \tag{4.10}$$

The first term in (4.10) is asymptotically bounded by the central limit theorem. The denominator in the second term converges to γ^2. The numerator in the second term can be written as

$$n^{-1/2}(b - c\beta)' \cdot n^{-1} \Sigma_i (x_i - \bar{x})'(x_i - \bar{x}) \cdot n^{-1/2}(b - c\beta) \quad . \tag{4.11}$$

The second term in (4.11) converges almost surely to $\ddagger$. The third term can be decomposed into

$$n^{-1/2}(b - \gamma\beta) - n^{-1/2}(c - \gamma) \cdot \beta \quad . \tag{4.12}$$

The first term is asymptotically bounded by theorem 4. The second term is asymptotically bounded by the central limit theorem applied to (3.11). Therefore (4.11) is $O_p(1)$, thus (3b) is satisfied. The moment conditions (3d) reduces to (4f) for the general transformation family. ∎

Remark 4.1. The moment condition (4f) is the same as the moment condition in Duan (1983) which considers the smearing estimate when the OLS is applied on the transformed scale $g^{-1}(y_i)$. Duan (1983) noted that if $|g'|$ is monotonic, we can replace the moment condition (4f) by

$$E[g'(c+\varepsilon)]^2 < \infty \quad \text{for all} \quad c \quad . \tag{4.13}$$

If the GLF is the power function $g(\eta) = \eta^q$, the moment condition (4.9) reduces to

$$E(c+\varepsilon)^{2(q-1)} < \infty \quad \text{for all} \quad c \quad ,$$

which is satisfied if $q > 1$ when the error term ε follows a normal distribution. Note that $q > 1$ implies that the linearizing transformation is the power transformation

$$g^{-1}(y) = y^{1/q} \quad ,$$

where the power parameter $1/q$ falls between zero and one.

If the GLF is exponential, $g(\eta) = \exp(\eta)$, i.e., the linearizing transformation g^{-1} is the logarithmic transformation, the moment condition (4.9) reduces to $E \exp(2\varepsilon) < \infty$, which is satisfied for the normal error distribution.

4.4. Smearing Estimate for Pointwise Slope Vector

The pointwise slope vector given by (1.15) can be estimated by differentiating the smearing estimate (1.12):

$$\nabla s(x) = c^{-1} b \cdot n^{-1} \Sigma_i \; g_1(b_0 + c^{-1} \cdot xb, e_i) \quad . \tag{4.14}$$

The second factor in (4.14) estimates the derivative $\tilde{v}'(x\beta)$ in (1.15).

Corollary 6. The smearing estimate (4.14) for the pointwise slope vector in a GLLM with a completely specified GLF is weakly consistent under the assumptions in theorem 1 and (3a) - (3d) in theorem 3, with g replaced by g_1 in (3c), and g_1, g_2 replaced by g_{11}, g_{12} in (4.6) and (4.7). ∎

Corollary 7. For a general transformation model with a completely specified GLF, the smearing estimate

$$\nabla s(x) = c^{-1} b \cdot n^{-1} \Sigma_i \; g'(b_0 + c^{-1} \cdot x b_1 + e_i) \tag{4.20}$$

is weakly consistent for the pointwise slope (1.15) under the assumptions in theorem 1, assumption (4a) in theorem 4, and assumptions (4e - f) in corollary 5, with g replaced by g' in (4e) and g' replaced by g'' in (4f). ∎

<u>**Remark 4.2**</u>. For the generalized linear model, there is no need to "smear" over the residuals to estimate the pointwise slopes; we can simply differentiate (4.2):

$$\nabla s(x) = c^{-1} b \cdot \tilde{\nu}'(b_0 + c^{-1} x b) \quad , \tag{4.21}$$

which is consistent if b_0 and c are consistent.

5. REFERENCES

Brillinger, D. R. (1983), A generalized linear model with "Gaussian" regressor variables, in Bickel, Doksum, and Hodges (eds.): "A festschrift for Erich L. Lehmann," pp. 97-114, Wadsworth Statistics/Probability Series, Wadsworth International Group. Belmont, CA.

Cacoullous, T. (1963), "Characterization of normality by constant regression of linear statistics on another linear statistic," Ann. Math. Stat., 38, 1894-1898.

Chung, Ching-Fan and Goldberger, Arthur S. (1984), Proportional projections in limited dependent variable models, Econometrica, 52, 2, 531-534.

Cox, D. R. and Snell, E. J. (1968): A general definition of residuals (with discussions), JRSS B, Vol. 30, 248-275.

Duan, N. (1983), Smearing estimate: a nonparametric transformation method, J. Amer. Statist. Assoc., 78, 605-610.

Efron, B. (1982), "Transformation theory: how normal is a family of distributions?" Ann. Stat., 10, 323-339.

Goldberger, Arthur S., (1981), Linear regression after selection, J. of Econometrics, 15, 357-366.

Haggstrom, G. W. (1983), Logistic regression and discriminant analysis by ordinary least squares, J. Bus. & Econ. Statis., 1, 3, 229-238.

Kagan, Linnik, and Rao (1973), Characterization problems in mathematical statistics, Wiley, New York.

Maddala, G. S. (1983), Limited-dependent and qualitative variables in econometrics, Cambridge University Press, Cambridge, U.K.

McCullagh, P. and Nelder, J. A. (1983), Generalized Linear Models, Chapman and Hall, New York.

Scheffé, H. (1959), Analysis of Variance," Wiley, New York.

Stein, C. M. (1981), Estimation of the mean of a multivariate normal distribution, Annals of Statistics, 9, 1135-1151.

APPENDIX A: PROOF OF THEOREM 1 UNDER SPHERICAL SYMMETRY

We need to show that

$$E^{x\beta}(x-\mu)\theta = 0 \tag{A.1}$$

where $\beta'\Sigma\theta = 0$. Let $\beta^* = (\beta'\Sigma\beta)^{-1/2} \cdot \beta$. For $p > 2$, we can find $\delta_2,\ldots,\delta_p$ such that $B = [\theta,\beta^*,\delta_2,\ldots,\delta_p]$ is unitary: $B'\Sigma B = I$.

Let $\xi = (x-\mu)B$. Since x is spherically symmetrical centered at μ, ξ has the same distribution as $(x - \mu)\Sigma^{-1/2}$. (Let $\xi^* = \xi\Sigma^{1/2} = (x-\mu)B\Sigma^{1/2}$; note that $(B\Sigma^{1/2})'\Sigma(B\Sigma^{1/2}) = \Sigma$, therefore ξ^* has the same distribution as $x-\mu$.)

<u>Claim</u> ξ is spherically symmetric with respect to the usual inner product $[v,w] = v'w$, i.e., ξA has the same distribution as ξ if $A'A = I$.

(Proof) Note that $A'B\Sigma BA = I$, thus BA is unitary with respect to the inner product $\langle\cdot,\cdot\rangle$. Therefore $\xi A = (x-\mu)BA$ has the same distribution as $x - \mu)\Sigma^{-1/2}$. ∎

Consider the diagonal matrix A with $A_{11} = -1$ and all other diagonal elements being one. It follows from the claim that

$$\xi A = \langle -\xi_1,\xi_2,\ldots,\xi_p\rangle$$

has the same distribution as ξ. Therefore

$$E(-\xi_1|\xi_2,\ldots,\xi_p) = E(\xi_1|\xi_2,\ldots,\xi_p) \equiv 0 \quad . \tag{A.2}$$

Integrating (A.2) with respect to $\xi_3,\ldots,\xi_p$, we have

$$E^{x\beta}(x-\mu)\theta = E(\xi_1|\xi_2) \equiv 0 \quad .$$

This proves (A.1) and completes the proof of theorem 1. ∎

APPENDIX B: PROOF OF THEOREM 4

Let $\delta = (b_0-\beta_0) + x(c^{-1}b-\beta)$, $\tau_i = e_i - \varepsilon_i$. Taking the first order Taylor's expansion for

$$\Delta_i = g(b_0+c^{-1}xb,\ e_i) - g(\beta_0+x\beta,\ \varepsilon_i)$$

in the direction (δ,τ_i), we have

$$\Delta_i = \delta\cdot g_{1i} + \tau_i\cdot g_{2i} \quad ,$$

where $0 < \theta_i < 1$, and $g_{ji} = g_j(\beta_0 + x\beta + \theta_i\delta,\ \varepsilon_i + \theta_i\tau_i)$, $j = 1,2$.

We need to show that

$$\delta\cdot n^{-1}\Sigma_i g_{1i} \to 0 \quad (p) \quad , \tag{B.1}$$

and

$$n^{-1}\Sigma_i\tau_i\cdot g_{2i} \to 0 \quad (p) \quad . \tag{B.2}$$

By assumption (4a) and theorem 1, the first factor in (B.1), δ, converges to zero in probability. By Cauchy-Schwarz inequality, the square of (B.2) can be bounded by

$$n^{-1}\Sigma_i\tau_i^2\cdot n^{-1}\Sigma_i(g_{2i})^2 \quad . \tag{B.3}$$

By assumption (4b), the first factor in (B.3) converges to zero in probability. It remains to show that the second factors in (B.1) and (B.3) are bounded asymptotically.

Since $\delta \to 0$ (p), for any $a > 0$, we can find n large enough so that $|\delta| < a$ with probability arbitrarily close to one. By assumption (4b), we can choose M large enough so that for n large enough, the inequality

$$\Sigma_i(e_i - \varepsilon_i)^2 < M^2$$

holds with probability arbitrarily close to one. In particular, we then have

$$|\tau_i| = |e_i - \varepsilon_i| < M, \quad i = 1,\ldots,n \quad .$$

The second factor in (B.1) is then bounded from above by

$$n^{-1}\Sigma_i\{\sup[|g_1(\beta_0+x\beta+t,\ \varepsilon_i+s)|;\ |s| \le M,\ |t| < a\} \quad ,$$

which converges to the expectation in (4.6) by the strong law of large numbers. Likewise, the second factor in (B.3) is bounded from above by an i.i.d. average which converges to the expectation in (4.7). ∎

APPENDIX C: PROOF OF THEOREM 5

Let $s = X'QX/n$, $c = X'QY/n$, $\theta = \Sigma_{xy}'$. We need to show that

$$n^{1/2}[s^{-1}c - \Sigma^{-1}\theta] = o_p(1) \quad .$$

It suffices to show that

$$n^{1/2}(c - \theta) = O_p(1) \quad , \tag{C.1}$$

$$n^{1/2}(s^{-1} - \Sigma^{-1}) = O_p(1) \quad . \tag{C.2}$$

The left hand side of (C.1) is equivalent to

$$\begin{aligned} &n^{-1/2}\Sigma_i[(x_i-\bar{x})y_i - E(x_i-\mu)y] \\ &= n^{-1/2}\Sigma_i[(x_i-\mu)y_i - E(x-\mu)y] + n^{1/2}(\bar{x}-\mu)\cdot\bar{y} \quad . \end{aligned} \tag{C.3}$$

By central limit theorem, the first term in (C.3) converges to a multinormal distribution under assumption (6a). The first factor in the second term converges to a multinormal distribution; the second factor converges to $E(y)$. Therefore (C.1) is satisfied.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER #2880	2. GOVT ACCESSION NO. AD-A163	3. RECIPIENT'S CATALOG NUMBER 599
4. TITLE (and Subtitle) The Ordinary Least Squares Estimation for the General-Link Linear Models, with Applications		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Naihua Duan and Ker-Chau Li		8. CONTRACT OR GRANT NUMBER(s) MCS 83-01587 DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53705		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS see item 18 below		12. REPORT DATE October 1985
		13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

general link function, spherical symmetry, polynomial normality, nonparametric regression, smearing estimate, linearized response surface, average slope vector, dichotomous regression, Tobit regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

For a general link linear model (GLLM), we show the the OLS estimate of the slope vector is strongly consistent up to a multiplicative scale, even though the model might actually be nonlinear. Furthermore, the estimated slope vector is strongly consistent for the average slope vector, the average of the pointwise slope vectors on the response surface. For a GLLM with a completely specified link function, we can solve for the multiplicative scalar and estimate the true slope vector, and estimate the intercept and Cox and Snell's generalized residuals. We then estimate the response surface and the pointwise slopes using

ABSTRACT (continued)

a generalization of the smearing estimate in Duan (1983). The results can be applied to a number of important subclasses of GLLM, including general transformation models, general scaled transformation models, generalized linear models, dichotomous regression, and Tobit regression.

END
FILMED
3-86
DTIC